Refinement Metrics for Quantitative Information Flow

Konstantinos Chatzikokolakis, Geoffrey Smith

To cite this version:

Refinement Metrics for Quantitative Information Flow

Konstantinos Chatzikokolakis[1] and Geoffrey Smith[2]

[1] University of Athens kostasc@di.uoa.gr
[2] Florida International University smithg@cis.fiu.edu

Abstract. In Quantitative Information Flow, *refinement* ($\sqsubseteq$) expresses the strong property that one channel never leaks more than another. Since two channels are then typically incomparable, here we explore a family of *refinement quasimetrics* offering greater flexibility. We show these quasimetrics let us unify refinement and capacity, we show that some of them can be computed efficiently via linear programming, and we establish upper bounds via the Earth Mover's distance. We illustrate our techniques on the Crowds protocol.

1 Introduction

Completely eliminating the leakage of sensitive information by computer systems is often infeasible, making it attractive to approach the problem quantitatively. *Quantitative information flow* [3] offers a rich family of *g-leakage* measures of the amount of leakage caused by a channel C taking a secret input X to an observable output Y; these measures are parameterized by X's *prior distribution* π and a *gain function* g, which models the adversary's capabilities and goals. (See §2 for a brief review.)

While g-leakage lets us precisely measure information leakage in a rich variety of operational scenarios, we may be unsure about the appropriate prior π or gain function g to use, and hence about the robustness of our conclusions. Two approaches to robustness have proved fruitful: *capacity*, which is the maximum leakage over some sets of gain functions and priors, and *refinement*: channel A is *refined* by channel B, written $\mathsf{A} \sqsubseteq \mathsf{B}$, iff B never leaks more than A, regardless of the prior or gain function. Remarkably, refinement also has a structural characterization: $\mathsf{A} \sqsubseteq \mathsf{B}$ iff there exists a "post-processing" channel matrix R such that $\mathsf{B} = \mathsf{AR}$. Moreover, $(\sqsubseteq)$ is a *partial order* on abstract channels.

Unfortunately, refinement is a very *partial* partial order, in that most channels are incomparable. And this "Boolean" nature of refinement is inconsistent with the spirit of quantitative information flow, which is above all motivated by the need to tolerate imperfect security. If $\mathsf{A} \not\sqsubseteq \mathsf{B}$, then we know that B can be worse than A; but we do not know whether B is really terrible, or whether is is just slightly worse than A. This is the main issue that we address in this paper.

In mathematics, it is common to use metrics to provide a "finer", quantified generalization of a relation. A relation has a "Boolean" nature: elements are either related or not. In the metric version, related elements have distance 0; but for non-related elements the metric tells you *how much* the relation is violated. For instance, the Euclidean distance can be seen as a quantified generalization of the equality relation $(=)$ on $\mathbb{R}$ — the distance $|x - y|$ tells us how much $x = y$ is violated; 0 means that it is not violated at all.[3] This approach is not of course limited to symmetric relations. For instance, the quasimetric

$$\mathbf{q}^{+}_{<}(x,y) \;\; := \;\; \max\{y-x,0\}\ ,$$

can be seen as a quantified version of $(\geq)$; it measures how much $(\geq)$ is violated, and is 0 iff $x \geq y$.

In this paper we define a family of *refinement quasimetrics* $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}$ to measure how much refinement is violated. Refinement is then the *kernel* (i.e. elements at distance 0) of these metrics: we have $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A},\mathsf{B}) = 0$ iff $\mathsf{A} \sqsubseteq \mathsf{B}$. Through this approach, we make the following contributions:

- We treat both additive and multiplicative leakage together via additive and multiplicative quasimetrics $\mathbf{q}^{+}_{<}$ and $\mathbf{q}^{\times}_{<}$. (The latter crucially takes the $\log$ of the ratio, instead of just the ratio.)
- We observe that the capacity of C can be expressed as $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathbb{1},\mathsf{C})$, where $\mathbb{1}$ is the perfect channel leaking nothing, giving us a unified way of looking at both refinement and capacity.
- In §4 we show that, for a fixed prior π, the additive refinement metric over all gain functions in $\mathbb{G}^{\updownarrow}\mathcal{X}$ can be computed in polynomial time via linear programming.
- In §5 we prove (again for a fixed π) that the refinement metric is bounded by the Earth Mover's distance between $[\pi{\triangleright}\mathsf{A}]$ and $[\pi{\triangleright}\mathsf{B}]$, allowing it to be approximated very efficiently. We also prove a bound on the metric over all priors.
- In §6 we give a case study of our techniques on the Crowds protocol.

2 Preliminaries

2.1 Quantitative Information Flow

In *quantitative information flow* [6, 7, 9] an adversary's prior knowledge about a secret input X drawn from a finite set $\mathcal{X}$ is modeled as a probability distribution π, and a probabilistic system taking input X to observable output Y is modeled as an information-theoretic channel matrix C giving the conditional probabilities $p(y|x)$. Assuming that the adversary knows C, then each output y allows the adversary to update her knowledge about X to a posterior distribution $p_{X|y}$, which moreover has probability $p(y)$ of occurring. As a result, the effect of C is to map each prior distribution π to a *distribution on posterior distributions*, called a *hyper-distribution* and denoted $[\pi{\triangleright}\mathsf{C}]$. For example, channel

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/2$	0	0
x_2	0	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

maps $\pi = (1/4, 1/2, 1/4)$ to hyper

$[\pi{\triangleright}\mathsf{C}]$	$1/4$	$1/3$	$7/24$	$1/8$
x_1	$1/2$	$3/8$	0	0
x_2	0	$3/8$	$6/7$	1
x_3	$1/2$	$1/4$	$1/7$	0

.

[3] Another example is bisimulation metrics for probabilistic properties. Because bisimulation is a strong property that can be broken by tiny modifications to transition probabilities, a variety of bisimulation metrics have been proposed, measuring how much bisimulation is violated.

This mapping is called the *abstract channel* denoted by C.[4]

It is natural to measure the "vulnerability" of X using functions V_g from probability distributions to reals, which are parameterized with a *gain function* $g : \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ that models the operational scenario; here $\mathcal{W}$ is the set of *actions* that the adversary can make, and $g(w,x)$ is the adversary's *gain* for doing w when the secret's actual value is x. We then define the g-*vulnerability* by $V_g(\pi) = \sup_w \sum_x \pi_x g(w,x)$, since that is the maximum expected gain over all possible actions. The choice of g allows us to model a wide variety of operational scenarios; a commonly used one is the *identity* gain function, given by $\mathcal{W} = \mathcal{X}$ and $g_{\text{id}}(w,x) = 1$ iff $w = x$ and 0 otherwise. For this gain function $V_{g_{\text{id}}}(\pi)$ gives the probability of correctly guessing the secret in one try.

Posterior vulnerability is defined as the average g-vulnerability in the hyper distribution: $V_g[\pi{\triangleright}\mathsf{C}] = \sum_y p(y) V_g(p_{X|y})$. And then g-*leakage* is defined as either the *ratio* or the *difference* between the posterior- and prior g-vulnerability; these are respectively *multiplicative leakage* $\mathcal{L}^{\times}_g(\pi, \mathsf{C})$ and *additive leakage* $\mathcal{L}^{+}_g(\pi, \mathsf{C})$.

2.2 Metrics

A *proper metric* on a set $\mathcal{A}$ is a function $d : \mathcal{A} \to [0, \infty]$ satisfying the following axioms for all $x, y, z \colon \mathcal{A}$:

1. **Reflexivity:** $d(x,x) = 0$.
2. **Symmetry:** $d(x,y) = d(y,x)$.
3. **Anti-symmetry:** $d(x,y) = d(y,x) = 0 \Rightarrow x = y$.[5]
4. **Triangle inequality:** $d(x,z) \leq d(x,y) + d(y,z)$.
5. **Finiteness**: $d(x,y) \neq \infty$.

Various prefixes can be added to "metric" to denote that some of the above conditions are dropped. The following are used in this paper:

1. **Quasi:** symmetry is dropped.
2. **Pseudo:** anti-symmetry is dropped.
3. **Extended:** finiteness is dropped.

Any combination of these prefixes is meaningful. For brevity, we use "metric" to generally refer to *any* of these classes, while we use the exact prefixes (or "proper") to refer to a precise one.

Kernel. The *kernel* of a metric d is a relation, containing points at distance 0:

$$(x,y) \in \ker(d) \quad \text{iff} \quad d(x,y) = 0\ .$$

It is easy to see that the metric properties of d imply the corresponding homonymous properties on relations for $\ker(d)$ (reflexivity, symmetry, anti-symmetry), while the triangle inequality of d corresponds to transitivity for $\ker(d)$. Hence the kernel of a proper

[4] Because of structural redundancies (e.g. the ordering and labels of columns), distinct channel matrices may denote the same abstract channel.

[5] When d is not symmetric, this is weaker than the axiom $d(x,y) = 0 \Rightarrow x = y$ (which is sometimes used in the literature even in the non-symmetric case).

metric is always the *equality relation* $=$, the kernel of a pseudometric is an *equivalence relation* (reflexive, symmetric and transitive), and the kernel of a quasimetric is a *partial order* (reflexive, anti-symmetric and transitive).

Quasimetrics on $\mathbb{R}$, $\mathbb{R}_{\geq 0}$. On $\mathbb{R}$ the following quasimetric is of particular interest:

$$\mathbf{q}^{+}_{<}(x,y) \quad := \quad \max\{y-x, 0\} \ .$$

Intuitively, $\mathbf{q}^{+}_{<}$ measures (additively) "how much smaller" than y is x; 0 means that x is "no smaller" than y. Or we can view $\mathbf{q}^{+}_{<}$ as measuring "how much $x \geq y$ is violated"; 0 means that $x \geq y$ holds (is not violated at all).

On $\mathbb{R}_{\geq 0}$ we can define a multiplicative variant of $\mathbf{q}^{+}_{<}$ as follows:[6]

$$\mathbf{q}^{\times}_{<}(x,y) \quad := \quad \max\{\log_2 \frac{y}{x}, 0\} \ ,$$

with the understanding that $\mathbf{q}^{\times}_{<}(0,y) = 0$ iff $y = 0$ and ∞ otherwise ($\mathbf{q}^{\times}_{<}$ is an extended quasimetric). Again, $\mathbf{q}^{\times}_{<}$ measures "how much smaller" than y is x, but this time multiplicatively. Although $\mathbf{q}^{+}_{<}$ and $\mathbf{q}^{\times}_{<}$ are clearly different, they coincide on their kernel:

$$\ker(\mathbf{q}^{+}_{<}) \quad = \quad \ker(\mathbf{q}^{\times}_{<}) \quad = \quad \geq \ .$$

Showing the quasimetric properties of $\mathbf{q}^{+}_{<}$ is trivial for all but the triangle inequality; for the latter, assuming $x \leq z$ (the case $x > z$ is trivial) we have that

$$\begin{array}{lll}
 & \mathbf{q}^{+}_{<}(x,z) & \\
= & z - x & \text{“}x \leq z\text{”} \\
= & z - y + y - x & \\
\leq & \max\{z-y, 0\} + \max\{y-x, 0\} & \\
= & \mathbf{q}^{+}_{<}(x,y) + \mathbf{q}^{+}_{<}(y,z) \ . &
\end{array}$$

Given a function $f : \mathcal{B} \to \mathcal{A}$ and metric d on $\mathcal{A}$ lets us define a metric $d \circ f$ on $\mathcal{B}$ by $(d \circ f)(x,y) := d(f(x), f(y))$.

Proposition 1. *The composition $d \circ f$ always preserves all metric properties of d except anti-symmetry, which is also preserved if f is injective.*

This construction is useful in that $\mathbf{q}^{\times}_{<} = \mathbf{q}^{+}_{<} \circ \log_2$, from which it follows that $\mathbf{q}^{\times}_{<}$ is an extended quasimetric on $\mathbb{R}_{\geq 0}$.[7]

3 Refinement metrics

To design metrics for QIF, we start from the simple observation that we can *compare vulnerabilities* using any *(quasi) metric q on* $\mathbb{R}_{\geq 0}$. For instance, *leakage* can be measured by comparing prior and posterior vulnerabilities as follows:

$$\mathcal{L}^{q}_{g}(\pi, C) \quad := \quad q(V_g(\pi), V_g[\pi \triangleright C]) \ .$$

[6] The use of logarithm is crucial for reflexivity and the triangle inequality. The choice of the base is arbitrary, but $\log_2$ is interesting due to the connection with min-entropy.

[7] Technically, to establish this we need to view $\mathbf{q}^{+}_{<}$ as an extended quasimetric on $\mathbb{R} \cup \{-\infty\}$ and $\log_2$ as an injective function $\mathbb{R}_{\geq 0} \to \mathbb{R} \cup \{-\infty\}$.

The usual *additive* and *multiplicative* variants of leakage can be obtained by instantiating q by $\mathbf{q}^{+}_{<}, \mathbf{q}^{\times}_{<}$ (except that for the latter we get $\log_2$ of the multiplicative leakage[8]):

$$\mathcal{L}_g^{\mathbf{q}^{+}_{<}}(\pi, C) \quad = \quad \mathcal{L}_g^{+}(\pi, \mathsf{C})\ , \qquad\qquad \mathcal{L}_g^{\mathbf{q}^{\times}_{<}}(\pi, C) \quad = \quad \log_2 \mathcal{L}_g^{\times}(\pi, \mathsf{C})\ .$$

The conceptual advantage of explicitly using a quasimetric (instead of just the difference or ratio) is twofold: first, we can exploit the metric properties of q and second, we can treat both leakage variants together.

Continuing this line of reasoning, we can obtain a *metric on channels*, by using q to *compare their posterior vulnerabilities*. This brings us to our refinement metric.

Definition 1. *Given classes $\mathcal{D} \subseteq \mathbb{D}\mathcal{X}$ of distributions and $\mathcal{G} \subseteq \mathbb{G}\mathcal{X}$ of gain functions, and metric q on $\mathbb{R}_{\geq 0}$, the refinement metric is defined as*

$$\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B}) \;\; := \;\; \sup_{\pi\colon \mathcal{D}, g\colon \mathcal{G}} q(V_g[\pi {\triangleright} \mathsf{A}], V_g[\pi {\triangleright} \mathsf{B}])\ .$$

We write $\mathrm{ref}^q_{\pi,\mathcal{G}}$ instead of $\mathrm{ref}^q_{\{\pi\},\mathcal{G}}$ when $\mathcal{D} = \{\pi\}$, and similarly for $\mathcal{G}$.

Any metric could be meaningful for q (for instance symmetric ones). But here we restrict to quasimetrics such that $\ker(q) = (\geq)$, which we call *order-respecting*. The standard choices of interest are $\mathbf{q}^{+}_{<}$ and $\mathbf{q}^{\times}_{<}$, giving additive and multiplicative comparison, for which we write the refinement metrics as $\mathrm{ref}^{+}_{\mathcal{D},\mathcal{G}}$ and $\mathrm{ref}^{\times}_{\mathcal{D},\mathcal{G}}$ respectively. (Note however that many results are independent of the choice of q.)

When $\mathcal{D}$ contains at least one point prior, or when $\mathcal{G}$ contains at least one constant g, then the $\max$ in the definition of $\mathbf{q}^{+}_{<}, \mathbf{q}^{\times}_{<}$ can be eliminated, giving

$$\begin{aligned} \mathrm{ref}^{+}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B}) \quad &= \quad \sup_{\pi\colon \mathcal{D}, g\colon \mathcal{G}} V_g[\pi {\triangleright} \mathsf{B}] - V_g[\pi {\triangleright} \mathsf{A}] \qquad\qquad \text{and} \\ \mathrm{ref}^{\times}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B}) \quad &= \quad \sup_{\pi\colon \mathcal{D}, g\colon \mathcal{G}} \log_2 \frac{V_g[\pi {\triangleright} \mathsf{B}]}{V_g[\pi {\triangleright} \mathsf{A}]}\ . \end{aligned}$$

Intuitively, $\mathrm{ref}^{+}_{\mathcal{D},\mathcal{G}}, \mathrm{ref}^{\times}_{\mathcal{D},\mathcal{G}}$ measure (additively or multiplicatively) *how robust* it is to replace A by B, with respect to the vulnerability of the system. A value of 0 means no risk —B is never worse than A (for $\mathcal{D}, \mathcal{G}$)— while a positive value means that there is risk, but it might be small. Note that refinement is *asymmetric* by design: replacing A by B is inherently different than replacing B by A.

For fixed π and g, we can obtain $\mathrm{ref}^q_{\pi,g}$ as the composition

$$\mathrm{ref}^q_{\pi,g} \quad = \quad q \circ V_g[\pi {\triangleright} \cdot]\ , \tag{1}$$

and then obtain $\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}$ as the $\sup$ of a family of quasimetrics

$$\mathrm{ref}^q_{\mathcal{D},\mathcal{G}} \quad = \quad \sup_{\pi\colon \mathcal{D}, g\colon \mathcal{G}} \mathrm{ref}^q_{\pi,g}\ . \tag{2}$$

This brings us to the following proposition.

[8] This is reminiscent of min-entropy; conceptually, however, we did not use $\log_2$ to convert vulnerabilities to entropies, we just used $\mathbf{q}^{\times}_{<}$ to compare vulnerabilities via a metric of multiplicative nature (the $\log$ conveniently turns ratios into a metric).

Proposition 2. $\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}$ *always inherits reflexivity, symmetry and the triangle inequality from q, but not necessarily anti-symmetry and finiteness.*

Proof. From (1) and Prop. 1 we get that $\mathrm{ref}^q_{\pi,g}$ inherits all metric properties of q except anti-symmetry ($V_g[\pi{\triangleright}\cdot]$ is not injective). Then, from (2), we only need to show that the $\sup$ of metrics preserves reflexivity, symmetry and the triangle inequality. The first two are trivial; for the triangle inequality, let $d = \sup_i d_i$, we have:

$$
\begin{array}{lll}
 & d(x,z) & \\
= & \sup_i d_i(x,z) & \\
\le & \sup_i(d_i(x,y) + d_i(y,z)) & \text{“triangle ineq. for } d_i\text{”} \\
\le & \sup_i d_i(x,y) + \sup_i d_i(y,z) & \text{“}\sup_i(a_i+b_i) \le \sup_i a_i + \sup_i b_i\text{”} \\
= & d(x,y) + d(y,z)\ . &
\end{array}
$$

Clearly, anti-symmetry is not always preserved (take $\mathcal{G}$ to contain only constant gain functions), nor is finiteness (for $\mathcal{G} = \mathbb{G}\mathcal{X}$ the $\sup$ can be ∞ even if q itself is finite). □

Hence $\mathrm{ref}^+_{\mathcal{D},\mathcal{G}}$ and $\mathrm{ref}^\times_{\mathcal{D},\mathcal{G}}$ always satisfy reflexivity and the triangle inequality (i.e. they are extended quasi pseudo metrics); they might satisfy additional properties for specific choices of $\mathcal{D}$ and $\mathcal{G}$.

3.1 Recovering the refinement relation

The following result states that if $\mathcal{D}$ and $\mathcal{G}$ are “sufficiently rich”, then we can obtain $\sqsubseteq$ as the kernel of the corresponding refinement metric $\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}$.

Theorem 1. *Assume that q is order-respecting, $\mathcal{D}$ contains a full-support prior, and $\mathcal{G}$ contains the non-negative, 1-bounded gain functions with finitely many actions. Then*

$$\ker(\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}) \quad = \quad \sqsubseteq\ .$$

Proof. (Sketch.) If $\mathsf{A} \not\sqsubseteq \mathsf{B}$, then the proof of the Coriaceous Theorem [10, Thm. 9] shows that for the uniform prior π^u, there is a gain function g, restricted as in the statement of the theorem, such that $V_g[\pi^u{\triangleright}\mathsf{A}] < V_g[\pi^u{\triangleright}\mathsf{B}]$; hence if q is order-respecting we have $\mathrm{ref}^q_{\pi^u,\mathcal{G}}(\mathsf{A},\mathsf{B}) > 0$. And by embedding the prior into g, we can replace π^u with any full-support prior. □

Hence the anti-symmetry of $\sqsubseteq$ on abstract channels is transferred to $\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}$:

Corollary 1. *Under the conditions of Thm. 1, $\mathrm{ref}^q_{\mathcal{D},\mathcal{G}}$ is anti-symmetric on abstract channels.*

3.2 Recovering capacity

Since the noninterfering channel $\mathbb{1}$ always produces a point hyper, capacity is simply given by the corresponding refinement metric between $\mathbb{1}$ and C.

Proposition 3. *For any q and $\mathcal{D},\mathcal{G}$ we have that $\mathcal{ML}^q_{\mathcal{G}}(\mathcal{D},\mathsf{C}) = \mathrm{ref}^q_{\mathcal{D},\mathcal{G}}(\mathbb{1},\mathsf{C})$.*

Proof. Direct consequence of the fact that $V_g[\pi \triangleright \mathbb{1}] = V_g(\pi)$. $\square$

Recall that using the quasimetrics $\mathrm{ref}^{+}_{\mathcal{D},\mathcal{G}}$ or $\mathrm{ref}^{\times}_{\mathcal{D},\mathcal{G}}$ for q make $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}$ itself non-symmetric. In particular $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathbb{1}, \mathsf{C})$ gives the capacity of C, while $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{C}, \mathbb{1})$ is always zero since $\mathsf{C} \sqsubseteq \mathbb{1}$.

A corollary of the above proposition is that the refinement metric of A, B can never exceed the capacity of B. This is intuitive since $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B})$ measures "how much worse than A can B be", while the capacity measures "how much worse than $\mathbb{1}$ (the best possible channel) can B be".

Corollary 2. *For any order-respecting q and $\mathcal{D}, \mathcal{G}, \mathsf{A}, \mathsf{B}$ we have that*

$$\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B}) \quad \leq \quad \mathcal{ML}^{q}_{\mathcal{G}}(\mathcal{D}, \mathsf{B}) \; .$$

Proof. We have that:

$$\begin{array}{lll}
 & \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathsf{B}) & \\
\leq & \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathbb{1}) + \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathbb{1}, \mathsf{B}) & \text{“triangle ineq. for } \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}\text{”} \\
\leq & \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathbb{1}, \mathsf{B}) & \text{“} \mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}(\mathsf{A}, \mathbb{1}) = 0 \text{ for order-respecting } q\text{”} \\
= & \mathcal{ML}^{q}_{\mathcal{G}}(\mathcal{D}, \mathsf{B}) & \text{“Prop. 3”}
\end{array}$$

$\square$

The above corollary in turn implies that $\mathrm{ref}^{q}_{\mathcal{D},\mathcal{G}}$ is finite whenever the capacity itself is finite. In particular $\mathrm{ref}^{+}_{\mathbb{D},\mathbb{G}^{\updownarrow}}$ and $\mathrm{ref}^{\times}_{\mathbb{D},\mathbb{G}^{+}}$ are both finite.

4 Computing the additive refinement metric

Given channel matrices $\mathsf{A}{:}\, \mathcal{X} {\rightarrow\!\!\triangleright} \mathcal{Y}$ and $\mathsf{B}{:}\, \mathcal{X} {\rightarrow\!\!\triangleright} \mathcal{Z}$ where $\mathsf{A} \not\sqsubseteq \mathsf{B}$, we know that sometimes B leaks more than A. To determine how much worse B can be, we wish to compute some version of the refinement metric: here we focus on the additive refinement metric $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\updownarrow}}(\mathsf{A}, \mathsf{B})$, which measures the maximum amount by which the additive leakage of B can exceed that of A on prior π and over gain functions in $\mathbb{G}^{\updownarrow}\mathcal{X}$, the class of 1-bounded gain functions.[9]

It is known that we can do this computation via linear programming [4], where the elements of G (the matrix representation of g) are variables. But here we refine the previously-used approach, enabling it to be done far more efficiently than was previously known. First, to ensure that G is in $\mathbb{G}^{\updownarrow}\mathcal{X}$, we just need to constrain its entries to be at most 1, and (to ensure that its gain values are non-negative) include a special action $\perp$ whose gain values are all 0. A second issue is that the number of possible actions of G is not fixed. But it suffices for G to have $|\mathcal{Y}| + |\mathcal{Z}| + 1$ actions, since this allows

[9] These are the gain functions whose gain values are at most 1 and which always produce non-negative vulnerabilities (even though some of their gain values can be negative). We need to restrict to bounded gain functions to get interesting results, since otherwise we could get arbitrarily large leakage differences simply by scaling up the gain values, making the additive refinement metric infinite.

a different action for each output of A and B, along with the special $\perp$ action. Now we recall the trace-based formulation of posterior vulnerability:

$$V_g[\pi \triangleright \mathsf{A}] = \max_{\mathsf{S}^\mathsf{A}} \mathbf{tr}(\mathsf{G}\mathsf{D}^\pi \mathsf{A}\mathsf{S}^\mathsf{A}) \quad \text{and} \quad V_g[\pi \triangleright \mathsf{B}] = \max_{\mathsf{S}^\mathsf{B}} \mathbf{tr}(\mathsf{G}\mathsf{D}^\pi \mathsf{B}\mathsf{S}^\mathsf{B})$$

where $\mathsf{S}^\mathsf{A}\colon \mathcal{Y} {\rightarrow\!\!\triangleright} \mathcal{W}$ and $\mathsf{S}^\mathsf{B}\colon \mathcal{Z} {\rightarrow\!\!\triangleright} \mathcal{W}$ are strategies mapping outputs of A and B to actions. While the previously-used approach was to try exponentially-many strategies S^A and S^B, we now argue that this is actually unnecessary. For we can assume without loss of generality that the first $|\mathcal{Y}|$ rows of G contain, in order, optimal actions for the columns of A; the next $|\mathcal{Z}|$ rows of G contain, in order, optimal actions for the columns of B; and finally the last row of G is all 0, for the $\perp$ action. Achieving this just requires reordering the rows of G and possibly duplicating some actions (since the same action might be optimal for more than one column). Once this is done, we can use a fixed S^A that maps column j of A to row j of G, and a fixed S^B that maps column k of B to row $k{+}|\mathcal{Y}|$ of G. Now we can solve a *single* linear programming problem for that S^A and S^B:[10]

> Choose G to maximize $\mathbf{tr}(\mathsf{G}\mathsf{D}^\pi \mathsf{B}\mathsf{S}^\mathsf{B}) - \mathbf{tr}(\mathsf{G}\mathsf{D}^\pi \mathsf{A}\mathsf{S}^\mathsf{A})$ subject to G having elements of at most 1 and a final all-zero row, and *opt* (S^A),

where *opt* (S^A) constrains S^A to be optimal for G. Remarkably, this means that a maximizing gain function G can be found in *polynomial time*.

Let us consider some examples. Consider the following channels:

A	y_1	y_2	y_3
x_1	0.1	0.2	0.7
x_2	0.3	0.5	0.2
x_3	0.6	0.1	0.3

B	y_1	y_2	y_3
x_1	0.101	0.200	0.699
x_2	0.298	0.501	0.201
x_3	0.600	0.099	0.301

where B is formed by slightly tweaking the probabilities in A. First, given π we can compute the additive capacity of A, $\mathcal{ML}^+_{\mathbb{G}\updownarrow}(\pi, \mathsf{A}) = \mathrm{ref}^+_{\pi,\mathbb{G}\updownarrow}(\mathbb{1}, \mathsf{A})$. As predicted in [5, Thm. 6], we find that for any full-support π, the capacity is 0.6, which is indeed 1 minus the sum of the column minimums of A; moreover, the gain function that realizes the capacity is indeed the complement of the π-reciprocal gain function. For instance, when $\pi = ({}^1\!/\!_3, {}^1\!/\!_2, {}^1\!/\!_6)$ it is

G	x_1	x_2	x_3
w_1	-2	1	1
w_2	1	-1	1
w_3	1	1	-5
$\perp$	0	0	0

Next, because B is so close to A, we expect that its additive refinement metric should be small. On the uniform prior $\pi^u = ({}^1\!/\!_3, {}^1\!/\!_3, {}^1\!/\!_3)$, we find indeed that $\mathrm{ref}^+_{\pi^u,\mathbb{G}\updownarrow}(\mathsf{A}, \mathsf{B}) =$

[10] Note that deciding whether $A \sqsubseteq B$ can also be done using a single linear program; but here we achieve more: when $A \not\sqsubseteq B$ we know that B can leak more than A, but we want to also compute *how much* more.

$^{1063}/_{255000} \approx 0.00416863$, realized on the gain function

G	x_1	x_2	x_3
w_1	$^{72}/_{85}$	$^{-88}/_{255}$	$^{8}/_{255}$
w_2	1	$^{-7}/_{3}$	1
w_3	1	1	-7
$\perp$	0	0	0

As another example, consider the following channels from [4]:

A	y_1	y_2	y_3	y_4
x_1	0.1	0.4	0.1	0.4
x_2	0.2	0.2	0.3	0.3
x_3	0.5	0.1	0.1	0.3

B	z_1	z_2	z_3
x_1	0.2	0.22	0.58
x_2	0.2	0.4	0.4
x_3	0.35	0.4	0.25

They are interesting in that $\mathsf{A} \not\sqsubseteq \mathsf{B}$, but it was historically difficult to find a gain function that makes B leak more than A. On a uniform prior we find that $\mathrm{ref}^{+}_{\pi^u,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B}) = {}^{103}/_{13320} \approx 0.00773273$, realized on the gain function

G	x_1	x_2	x_3
w_1	$^{-107}/_{148}$	$^{409}/_{444}$	$^{19}/_{444}$
w_2	$^{-1}/_{4}$	1	$^{-2}/_{3}$
w_3	1	$^{-7}/_{3}$	1
$\perp$	0	0	0

To achieve robustness wrt any prior, we might prefer not to compute $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})$ for some specific π, but instead to compute $\mathrm{ref}^{+}_{\mathbb{D},\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})$. But it is not clear how to do this.

To get some partial insight, we can try an exhaustive search over all priors whose probabilities have a denominator of 1000. We find that the best prior in that set is $\pi = ({}^{624}/_{1000}, {}^{29}/_{1000}, {}^{347}/_{1000})$, which gives $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B}) = {}^{15553}/_{1850000} \approx 0.00840703$, realized on gain function

G	x_1	x_2	x_3
w_1	$^{-5}/_{7696}$	$^{16}/_{1073}$	$^{4}/_{12839}$
w_2	$^{57}/_{208}$	1	$^{-257}/_{347}$
w_3	1	$^{-1179}/_{29}$	1
$\perp$	0	0	0

Hence we see that unlike the situation of additive capacity where $\mathcal{ML}^{+}_{\mathbb{G}^{\ddagger}}(\pi,\mathsf{A})$ is the same for any full support π, the additive refinement metric *does* depend on the prior.[11]

[11] There is an interesting phenomenon that we have observed. Given prior π, we can first compute the gain function g that realizes $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})$, and then we can compute the prior that realizes $\mathrm{ref}^{+}_{\mathbb{D},g}(\mathsf{A},\mathsf{B})$. In every case that we have tried, the prior that is found is exactly the π that we started with, suggesting that π is somehow "encoded" into the gain function g.

5 Bounding the additive refinement metric

In the previous section we showed that computing the additive refinement metric for a fixed prior can be done by solving a *single* linear program. Although this gives us a solution in time polynomial in the size of C, this solution is still practically feasible only for very modest sizes. In this section, we study efficient techniques for bounding the additive metric based on the Earth Mover's distance, inspired by the use of the same technique for computing capacity [2].

Moreover, we discuss a simple bound on the additive refinement metric when maximizing over both priors and gain functions. The existence of an efficient exact algorithm for this case remains unknown.

5.1 The Earth Mover's distance

The *Earth Mover's* distance (EMD), also known as *Wasserstein* distance, is a fundamental distance between *probability distributions*, with numerous applications in computer science. EMD gives a metric on the set $\mathbb{D}\mathcal{A}$ of probability distributions over an underlying set $\mathcal{A}$, equipped with its own *"ground" metric* d. As a consequence, it can be thought of as a mapping $\mathbb{W} : \mathbb{M}\mathcal{A} \to \mathbb{M}\mathbb{D}\mathcal{A}$, lifting a metric on $\mathcal{A}$ to a metric on $\mathbb{D}\mathcal{A}$.

The idea behind EMD is that the distance between two distributions $\alpha, \alpha'{:}\,\mathbb{D}\mathcal{A}$ is the minimum *cost* of *transforming* α into α'. The transformation is performed by moving probability mass between elements $a, a'{:}\,\mathcal{A}$, where α_a is seen as the probability mass *available* at a (supply), and $\alpha'_{a'}$ as the probability mass *needed* at a' (demand). The ground metric d determines the *cost* of this transportation: moving a unit of probability from a to a' costs $d(a, a')$. There are many strategies for achieving this transformation, the cost of the best one gives the distance between α and α'.

More precisely, an earth-moving *strategy* is a joint distribution $S \in \mathbb{D}\mathcal{A}^2$ whose two marginals are α and α', i.e. $\sum_{a'} S(a, a') = \alpha_a$ and $\sum_a S_{a,a'} = \alpha'_{a'}$. Intuitively, $S_{a,a'}$ is the probability mass moved from a to a', and the requirements on the marginals capture the fact that both supply and demand should be respected. We write $\mathcal{S}_{\alpha,\alpha'}$ for the set of such strategies. Moreover, we write $\mathcal{E}_S d{:=} \sum_{a,a'} S_{a,a'} d(a, a')$ for the expected value of d wrt the distribution S, in other words the total transportation cost of the strategy S.

Definition 2. *The Earth mover's distance is the mapping* $\mathbb{W} : \mathbb{M}\mathcal{A} \to \mathbb{M}\mathbb{D}\mathcal{A}$ *given by:*

$$\mathbb{W}(d)(\alpha, \alpha') \quad := \quad \inf_{S{:}\,\mathcal{S}_{\alpha,\alpha'}} \mathcal{E}_S d \; .$$

Note that the properties of $\mathbb{W}(d)$ directly depend on those of d; in particular $\mathbb{W}(d)$ might not be symmetric unless d itself is symmetric.

The EMD can be computed via a linear program with S as variables, since $\mathcal{E}_S d$ is linear on S and respecting supply and demand can be expressed by linear constraints. However, more efficient network flow algorithms also exist (e.g. [11]). Note, however, that some of these algorithms require d to be a proper metric.

5.2 Bounding the refinement metric using the EMD

We now turn our attention to the problem of bounding the additive refinement metric between two channels A and B. To do so, we exploit the view of channels as producing *hyper-distributions*. Recall from §2 that the effect of a channel A on the prior π can be viewed as a hyper-distribution $[\pi{\triangleright}\mathsf{A}]$, that is a distribution on the posterior distributions produced by each output of A. In other words, hypers have type $\mathbb{D}\mathbb{D}\mathcal{X}$.

Since hypers are distributions, we can use the EMD to measure the distance between them. But the EMD requires a ground metric on the underlying space, which in our case is $\mathcal{A} = \mathbb{D}\mathcal{X}$. In other words, to measure the distance between hypers we first need to measure the distance between posteriors. It turns out that the appropriate distance for our needs is the following.

Definition 3. *Let $\pi, \sigma{:}\,\mathbb{D}\mathcal{X}$. Define the "convex separation" quasimetric $\mathbf{q_{cs}}{:}\,\mathbb{M}\mathbb{D}\mathcal{X}$ as*

$$\mathbf{q_{cs}}(\sigma, \pi) \quad := \quad \max_{x{:}\,\lceil\sigma\rceil} \left(1 - \frac{\pi_x}{\sigma_x}\right) \ .$$

The usefulness of this quasimetric lies in the fact that it can be shown [5] to express the maximum that σ, π can be separated by a 1-bounded gain function, that is

$$\mathbf{q_{cs}}(\sigma, \pi) = \sup_{g{:}\,\mathbb{G}^{\ddagger}\mathcal{X}} \mathbf{q}^{+}_{<}(V_g(\sigma), V_g(\pi)) \ .$$

This in turn means that V_g is non-expansive wrt $\mathbf{q_{cs}}$, which is important due to the connection between EMD and the Kantorovich metric. The technicalities of this construction are available in the appendix, we only state the most relevant result here.

Theorem 2. *For any prior π and channels A, B it holds that*

$$\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{A}, \mathsf{B}) \quad \le \quad \mathbb{W}(\mathbf{q_{cs}})([\pi{\triangleright}\mathsf{A}], [\pi{\triangleright}\mathsf{B}]) \ .$$

Hence, to obtain a bound on the refinement metric, we can compute the hypers $[\pi{\triangleright}\mathsf{A}]$ and $[\pi{\triangleright}\mathsf{B}]$ (i.e. the corresponding output and posterior distributions) and then employ either linear programming or a network flow algorithm to compute the $\mathbb{W}(\mathbf{q_{cs}})$ distance between them. Note that the algorithm for EMD should not require the ground metric to be symmetric, since $\mathbf{q_{cs}}$ is not.

5.3 More relaxed bounds

As discussed in the previous sections, the EMD is the cost of the *best* transportation strategy to transform one distribution into another. As a consequence, one can clearly obtain a bound by choosing *any* transportation strategy, not necessarily the best. That is, if S is a strategy for transforming $[\pi{\triangleright}\mathsf{A}]$ into $[\pi{\triangleright}\mathsf{B}]$, then the cost (wrt $\mathbf{q_{cs}}$) of S provides a bound to the refinement metric between A and B:

$$\mathrm{ref}^{+}_{\pi,\mathcal{G}}(\mathsf{A}, \mathsf{B}) \quad \le \quad \mathcal{E}_S \mathbf{q_{cs}} \ .$$

Of course the actual EMD gives the best such bound, but the freedom of choosing *any* strategy can be convenient. In §6 an analytic bound for the Crowds protocol is given by choosing an arbitrary strategy, without having to show that this is the optimal one.

5.4 Maximizing over all priors

In view of the apparent difficulty of computing $\mathrm{ref}^{+}_{\mathbb{D},\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})$, we are led to wonder how much the additive refinement bound can be increased by changing from a uniform prior to a non-uniform prior. We find that the possible increase can be bounded:

Theorem 3. *For any channel matrices* $\mathsf{A}\colon \mathcal{X} {\to\!\!\triangleright} \mathcal{Y}$ *and* $\mathsf{B}\colon \mathcal{X} {\to\!\!\triangleright} \mathcal{Z}$*, we have*

$$\mathrm{ref}^{+}_{\mathbb{D},\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B}) \quad\leq\quad |\mathcal{X}|\cdot \mathrm{ref}^{+}_{\pi^u,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})\ .$$

Proof. Let prior π be arbitrary, and suppose that $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{A},\mathsf{B})$ is realized on gain function g with matrix representation G, which we assume without loss of generality to include an all-zero row. Next recall the trace-based formulation of posterior g-vulnerability: $V_g[\pi{\triangleright}\mathsf{A}] = \max_{\mathsf{S}} \mathbf{tr}(\mathsf{G}\mathsf{D}^{\pi}\mathsf{A}\mathsf{S})$. Observe that we can factor $\mathsf{D}^{\pi} = \mathsf{D}^{\rho}\mathsf{D}^{\pi^u}$, where vector ρ is given by $\rho_x = {\pi_x}/{\pi^u}_x = |\mathcal{X}|\cdot\pi_x$. Now we can translate from π to the uniform prior π^u:

$$V_g[\pi{\triangleright}\mathsf{A}] \quad=\quad \max_{\mathsf{S}} \mathbf{tr}(\mathsf{G}\mathsf{D}^{\pi}\mathsf{A}\mathsf{S}) \quad=\quad \max_{\mathsf{S}} \mathbf{tr}(\mathsf{G}\mathsf{D}^{\rho}\mathsf{D}^{\pi^u}\mathsf{A}\mathsf{S}) \quad=\quad V_{g'}[\pi^u{\triangleright}\mathsf{A}]\ ,$$

where g' is the gain function with matrix representation $\mathsf{G}\mathsf{D}^{\rho}$. The only problem is that g' may not be in $\mathbb{G}^{\ddagger}\mathcal{X}$. It *does* give non-negative vulnerabilities (since $\mathsf{G}\mathsf{D}^{\rho}$ contains an all-zero row), but it may have gain values bigger than 1. We can deal with this by scaling g' down, dividing its gain values by $\max_x {\pi_x}/{\pi^u}_x = \max_x \pi_x\cdot|\mathcal{X}|$, which is at most $|\mathcal{X}|$. If we let g'' be the scaled-down version of g', then we have $g''\in\mathbb{G}^{\ddagger}\mathcal{X}$ and $|\mathcal{X}|\cdot V_{g''}[\pi^u{\triangleright}\mathsf{A}] = V_g[\pi{\triangleright}\mathsf{A}]$. Since the same equality holds for B, the desired inequality follows. (Note that it is only an inequality, since g'' may not be optimal for π^u.) $\square$

Applying this theorem to the last example from §4, where the additive refinement metric on the uniform prior was about 0.0077, we see that the metric over *all* priors is at most 3 times as large, or 0.0232. While this bound is considerably larger than 0.0084 (the maximum found in our partial search), it may still be useful.

6 Case study : the Crowds protocol

Crowds is a simple protocol for anonymous web surfing. The context is that a user, called the *initiator*, wants to contact a *web server* but does not want to disclose his identity to the server. That is achieved by collaborating with a group of other users, called the *crowd*, who participate in the protocol to facilitate the task. The protocol is essentially a simple probabilistic routing protocol:

- In the first step, the initiator selects a user uniformly (including possibly himself) and forwards the request to him. The user who receives the message now becomes the (new) *forwarder*.
- A forwarder, upon receiving a message, flips a (biased) probabilistic coin: with probability φ he forwards the message to a new user (again chosen uniformly, including himself), but with remaining probability $1{-}\varphi$ he instead delivers the message directly to the web server.

If $\varphi{<}1$ then with probability 1 the message will eventually arrive at the web server, with the expected number of hops being $1/1{-}\varphi$.

Concerning anonymity, from the point of view of the web server all users seem equally likely to have been the initiator (assuming a uniform prior), due to the first step's always being a forwarding step. The analysis becomes more interesting however if we assume that some users in the crowd are corrupt, reporting to an adversary who is trying to discover messages' initiators.[12]

If user a forwards a message to a corrupted user, we say that a was *detected*. Since a's being detected is more likely when he is the initiator than when he is not, the protocol does leak information: detecting a creates a posterior where a has a greater chance of being the initiator than he had *a priori*. Still, if $\varphi{>}0$ then the posterior is *not* a point distribution: user a can "plead not guilty" by claiming the he was merely forwarding for someone else. The greater φ is, the more plausible that claim becomes: the forwarding probability φ can be seen as trading anonymity for utility (expected number of hops).

In their original work, Reiter and Rubin [12] introduced the property of "probable innocence" to analyze Crowds, meaning, roughly speaking, that each user appears more likely to be innocent than guilty. For our purposes, however, we study Crowds in the context of QIF.[13]

6.1 Modeling the Crowds protocol

To perform an analysis of the anonymity guarantees of Crowds, we start by modeling the protocol as an information-theoretic channel. Letting n, c, m be the number of honest, corrupted, and total users respectively (i.e. $n{+}c = m$), the secret and observable events are $\mathcal{X}{:=}\{x_1,\ldots,x_n\}$, $\mathcal{Y}{:=}\{y_1,\ldots,y_n,s\}$, with x_a denoting that (honest) user a is the initiator, y_b denoting that user b was detected (forwarded a message to a corrupted user), and s denoting that the message was delivered to the web server without having been forwarded to a corrupted user.

To construct the channel matrix C, we need for any honest user a to compute the probability of producing each observable event, given the secret event x_a. Due to symmetry, the probability of detecting the initiator, or some non-initiator, is the same for all initiators. In other words, the probabilities $p(y_b|x_a)$ only depend on whether $a{=}b$ or $a{\neq}b$, and not on the exact value of a,b. Hence the entries of C can only have 3 distinct values, $\mathsf{C}_{x_a,s} = \alpha$, $\mathsf{C}_{x_a,y_a} = \beta$, and $\mathsf{C}_{x_a,y_b} = \gamma$, $a \neq b$, giving the following channel matrix:[14]

$$\mathsf{C} \;=\; \begin{array}{c} \\ x_1 \\ \vdots \\ x_n \end{array}\!\!\begin{array}{c} \begin{array}{ccc} y_1 \cdots y_n & & s \end{array} \\ \left[\begin{array}{ccc} \beta \;\cdots\; \gamma & & \alpha \\ \ddots & & \\ \gamma \;\cdots\; \beta & & \alpha \end{array}\right] \end{array} \qquad \text{where} \qquad \begin{array}{rcl} \alpha & := & \dfrac{n-\varphi n}{m-\varphi n}\,, \\ \beta & := & \dfrac{c(m-\varphi(n{-}1))}{m(m-\varphi n)}\,, \\ \gamma & := & \dfrac{c\varphi}{m(m-\varphi n)}\,. \end{array}$$

[12] This is a relatively weak adversarial model; a model in which the adversary can view any communication in the network would be too strong for Crowds.

[13] See [3] for a discussion of how probable innocence relates to QIF.

[14] See [3] for the exact derivations.

Finally, given a prior π on $\mathcal{X}$, modeling how likely each user is to be the initiator a-priori, we can quantify the anonymity of the protocol by considering its (additive or multiplicative) g-leakage, for a properly constructed gain function g. For instance, the identity gain function g_{id} is a reasonable choice for this protocol, modeling an adversary that tries to guess the initiator of the message in one try. For this choice of gain function, and the uniform prior π^u, the additive leakage is given by

$$\mathcal{L}^{+}_{g_{\text{id}}}(\pi^u, \mathsf{C}) \quad = \quad \frac{nc - c}{nc + n} \quad . \tag{3}$$

Bt g_{id} is not the only reasonable choice for this protocol. It could be argued, for instance, that the whole point behind "probable innocence" is that the adversary would not incriminate a user unless there is sufficient evidence — that is, a *wrong* guess should have a *penalty*, compared to *not guessing at all*. Such adversaries can be expressed by suitable gain functions [3].

A crucial question, then, is how can we make *robust* conclusions in our analysis of information leakage, in particular when we tune a parameter of the protocol such as φ?

6.2 Increasing φ : refinement relation

A striking fact about Crowds, first observed in [8], is that its Bayes leakage for a uniform prior, given by (3), does not depend on φ! This is remarkable since the fact of forwarding is the only reason why a detected user can plausibly pretend not be the initiator. Intuitively, the more likely a user is to forward the more likely detecting the wrong user becomes, so anonymity is improved. Understanding this paradox is beyond the scope of this paper (an explanation is given in [3]), but we should note that this phenomenon *only* happens in the very specific case of g_{id} with a uniform prior: for most other priors and gain functions, the corresponding leakage does depend on φ, as expected.

This brings us to the natural question: *how* does leakage depend on φ? Answering this question is of great importance to the designer of the system who tries to optimize the parameters, but without compromising its anonymity.

The easier case to answer is when φ is *increased*. Intuitively, this operation should be *safe*, in the sense that the resulting protocol should leak no more than before. But can we be sure that this *always* happens? Could it be the case that for some strange adversary, modeled by some gain function g, increasing φ causes the leakage to increase, leading to a less safe protocol?

The theory of refinement provides a robust answer. Let C^{φ} denote the channel matrix for a particular choice of φ. It can be shown that

$$\mathsf{C}^{\varphi} \sqsubseteq \mathsf{C}^{\widetilde{\varphi}} \qquad \text{iff} \qquad \varphi \leq \widetilde{\varphi}\,,$$

by showing that $\mathsf{C}^{\widetilde{\varphi}} = \mathsf{C}^{\varphi}\mathsf{R}$ for a properly constructed channel R. As a consequence $\mathsf{C}^{\widetilde{\varphi}}$ is a safer protocol; it leaks no more than C^{φ} for *all* priors and gain functions.

6.3 Decreasing φ : refinement metric

Although refinement establishes in a very robust way that increasing φ is a safe operation, it can be argued that the interesting case is exactly the opposite: *decreasing* φ.

This is because a system designer is typically interested in improving the *utility* of the system, which in this case corresponds to its *latency*. Since the expected number of hops to reach the server is ${}^1/_{1-\varphi}$, decreasing φ lowers the protocol's latency, so such a change is desirable.

However, decreasing φ is *not always* safe. Although Bayes leakage for a uniform prior remains unaffected, we can experimentally verify that for all other priors, and for many gain functions, the leakage indeed increases. We thus have a typical trade-off between anonymity and utility; we might be willing to accept a less safe protocol, and *replace* $\widetilde{\varphi}$ *by a smaller* φ, if we knew that the increase in leakage cannot exceed a certain threshold. Can we answer this question in a *robust* way?

Using the refinement metric. The refinement metric provides a natural answer to our question. For a fixed π, we know that $\mathrm{ref}^{+}_{\pi,\mathbb{G}^{\ddagger}}(\mathsf{C}^{\widetilde{\varphi}}, \mathsf{C}^{\varphi})$ gives the maximum increase in vulnerability when we replace $\widetilde{\varphi}$ by φ, for any 1-bounded gain function $g{:}\,\mathbb{G}^{\ddagger}$.

The metric can be computed using the linear programming technique of §4). As an example, for $n = 110$, $c = 5$, $\varphi = 0.55$ and $\widetilde{\varphi} = 0.6$ we find that the refinement metric is approximately 0.0076. This means that lowering the probability of forwarding from 0.6 to 0.55 cannot have a huge effect on the vulnerability; for instance if the adversary's probability of a correct guess (Bayes vulnerability) was p, it can be at most $p + 0.0076$ after the change. Note that $\varphi = 0.55$ gives 12.5% *lower latency* compared to $\varphi = 0.6$, so we might decide that the benefit outweighs the risk.

If computing the exact metric is not feasible, we might still be able to apply the EMD bound discussed in §5, given by the EMD (wrt $\mathbf{q_{cs}}$) between the hypers $[\pi^u{\triangleright}\mathsf{C}^{\widetilde{\varphi}}]$ and $[\pi^u{\triangleright}\mathsf{C}^{\varphi}]$. In our example, this is approximately 0.017; it is more than twice as large as the real metric, but it still provides a useful bound on the loss of anonymity.

An analytic bound for Crowds. As discussed in §5, an upper bound to the refinement metric can be obtained by computing, not the actual EMD, but the cost (wrt $\mathbf{q_{cs}}$) of *any transportation strategy* from $[\pi{\triangleright}\mathsf{C}^{\widetilde{\varphi}}]$ to $[\pi{\triangleright}\mathsf{C}^{\varphi}]$. For the uniform prior, $[\pi^u{\triangleright}\mathsf{C}^{\varphi}]$ can be computed analytically; then we can choose any strategy to obtain an analytic bound.

Following the calculations of [3], the hyper distribution $[\pi^u{\triangleright}\mathsf{C}^{\varphi}]$ has

$$
\begin{array}{lrcl}
\text{output distribution} & \rho & = & ({}^{1-\alpha}/_{n}, \cdots, {}^{1-\alpha}/_{n}, \alpha)\ , \\
\text{and posteriors} & \delta^{y_b} & = & (\dfrac{\varphi}{m}, \cdots, \dfrac{\varphi}{m}, \dfrac{m-\varphi(n{-}1)}{m}, \dfrac{\varphi}{m}, \cdots, \dfrac{\varphi}{m})\ , \\
 & \delta^{s} & = & ({}^{1}/_{n}, \cdots, {}^{1}/_{n})\ .
\end{array}
$$

Note that the detection events y_b produce a posterior δ^{y_b} assigning higher probability to the detected user b, while s provides no information, producing a uniform posterior. Denoting by $\widetilde{\alpha}, \widetilde{\rho}, \ldots$ the corresponding quantities for $\widetilde{\varphi} \geq \varphi$, we can check that $\widetilde{\alpha} \leq \alpha$; hence $\widetilde{\rho}$ has lower probability than ρ on its last element and higher on its first n elements. A reasonable transportation strategy[15] S from $[\pi{\triangleright}\mathsf{C}^{\widetilde{\varphi}}]$ to $[\pi{\triangleright}\mathsf{C}^{\varphi}]$ is

[15] Experiments suggest that S might be optimal, although this is neither required nor proved.

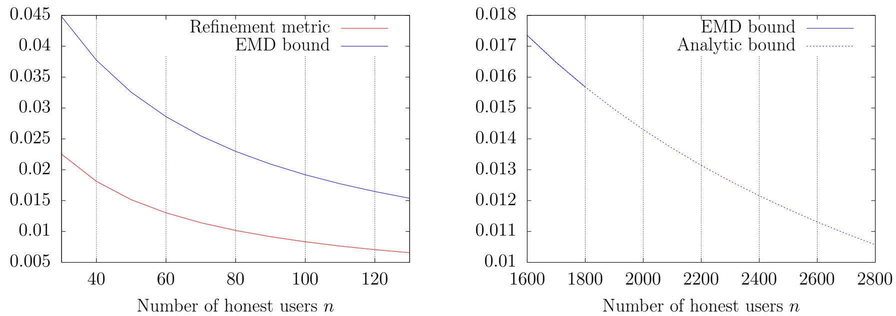

Fig. 1. Refinement metric, EMD and analytic bounds for Crowds

then to move all probability mass from $\widetilde{\delta}^s$ to δ^s, while splitting the probability of $\widetilde{\delta}^{y_b}$ between δ^{y_b} and δ^s. This strategy produces the following bound.

Theorem 4. *For any* $0 < \varphi \le \widetilde{\varphi} \le 1$ *it holds that:*

$$\mathrm{ref}^+_{\pi^u,\mathbb{G}\ddagger}(\mathsf{C}^{\widetilde{\varphi}}, \mathsf{C}^{\varphi}) \quad \le \quad (1-\alpha)(1-\frac{\varphi}{\widetilde{\varphi}}) + (\alpha - \alpha')\left(1 - \frac{m}{mn - \widetilde{\varphi}n(n-1)}\right) .$$

Experimental evaluation All techniques presented in this paper have been implemented in the libqif library [1]. An experimental evaluation of the additive refinement metric for Crowds is shown in Fig. 1. On the left-hand side, we compute the exact metric as well as the EMD bound between two instances of Crowds with $\varphi = 0.55$ and $\widetilde{\varphi} = 0.6$, for a uniform prior, $c = 5$ corrupted users, while varying the number of honest users n. We see that as n increases $\mathrm{ref}^+_{\pi^u,\mathbb{G}\ddagger}(\mathsf{C}^{\widetilde{\varphi}}, \mathsf{C}^{\varphi})$ becomes small, guaranteeing that the decrease of φ has limited effect on the leakage. The EMD bound is not tight, but as n increases it becomes small enough to provide meaningful guarantees.

Computing the exact metric was possible for up to $n = 130$ within a computation threshold of 10 minutes. Computing the EMD bound, on the other hand, was feasible for much larger sizes. On the right-hand side of Fig. 1 we show both the EMD and the analytic bound of the previous section, for the same example, but for larger numbers of users and $c = 70$. The two bounds coincide, suggesting that the transportation strategy chosen for the analytic bound is likely the optimal one. We were able to compute the EMD bound for up to $n = 1800$, while the analytic bound is applicable to any size. Note that, although the exact refinement metric is unknown, the bounds are sufficiently small to provide meaningful guarantees about the anonymity of the protocol.

References

1. libqif : Quantitative information flow c++ library. https://github.com/chatziko/libqif
2. Alvim, M.S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., Smith, G.: Additive and multiplicative notions of leakage, and their capacities. In: Proc. 27th IEEE Computer Security Foundations Symposium (CSF 2014). pp. 308–322 (2014)

3. Alvim, M.S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., Smith, G.: The Science of Quantitative Information Flow. Springer (2019)
4. Alvim, M.S., Chatzikokolakis, K., Palamidessi, C., Smith, G.: Measuring information leakage using generalized gain functions. In: Proc. 25th IEEE Computer Security Foundations Symposium (CSF 2012). pp. 265–279 (Jun 2012)
5. Chatzikokolakis, K.: On the additive capacity problem for quantitative information flow. In: Proc. 15th International Conference on Quantitative Evaluation of Systems (QEST). pp. 1–19 (2018)
6. Clark, D., Hunt, S., Malacaria, P.: Quantitative analysis of the leakage of confidential data. In: Proc. Workshop on Quantitative Aspects of Programming Languages. Electr. Notes Theor. Comput. Sci, vol. 59 (3), pp. 238–251 (2001)
7. Clarkson, M., Myers, A., Schneider, F.: Belief in information flow. In: Proc. 18th IEEE Computer Security Foundations Workshop (CSFW '05). pp. 31–45 (2005)
8. Espinoza, B., Smith, G.: Min-entropy as a resource. Information and Computation (Special Issue on Information Security as a Resource) 226, 57–75 (Apr 2013)
9. Köpf, B., Basin, D.: An information-theoretic model for adaptive side-channel attacks. In: Proc. 14th ACM Conference on Computer and Communications Security (CCS '07). pp. 286–296 (2007)
10. McIver, A., Morgan, C., Smith, G., Espinoza, B., Meinicke, L.: Abstract channels and their robust information-leakage ordering. In: Proc. 3rd Conference on Principles of Security and Trust (POST 2014). pp. 83–102 (2014)
11. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: ICCV (2009)
12. Reiter, M.K., Rubin, A.D.: Crowds: Anonymity for web transactions. ACM Transactions on Information Systems Security 1(1), 66–92 (1998)
13. Villani, C.: Topics in optimal transportation. No. 58, American Mathematical Soc. (2003)

A Proofs of Section 5

In this section we provide the proof of the main result of Section 5.

A.1 The Kantorovich distance

The Kantorovich metric plays a fundamental role in our bounding technique. Given metrics $d_{\mathcal{A}} : \mathbb{M}\mathcal{A}, d_{\mathcal{B}} \in \mathbb{M}\mathcal{B}$, a function $f : \mathcal{A} \to \mathcal{B}$ is called Lipschitz wrt $d_{\mathcal{A}}, d_{\mathcal{B}}$ iff

$$d_{\mathcal{B}}(f(a), f(a')) \quad \leq \quad d_{\mathcal{A}}(a, a') \qquad \text{for all } a, a' \in \mathcal{A} \ .$$

Let $\mathbb{C}^{d_{\mathcal{A}}, d_{\mathcal{B}}}\mathcal{A}$ denote the set of all such functions.

The standard Kantorovich construction transforms metrics on some set $\mathcal{A}$, to metrics on $\mathbb{D}\mathcal{X}$, i.e. metrics measuring the distance between two probability distributions on $\mathcal{A}$. Formally, it is the lifting $\mathbb{K} : \mathbb{M}\mathcal{A} \to \mathbb{M}\mathbb{D}\mathcal{A}$ given by

$$\mathbb{K}(d)(\alpha, \alpha') \ \ := \ \ \sup_{F \colon \mathbb{C}^{d, d_{\mathbb{R}}}\mathcal{A}} |\mathcal{E}_{\alpha} F - \mathcal{E}_{\alpha'} F| \ .$$

Note the two uses of the Euclidean distance $d_{\mathbb{R}}$ in the above definition: first, we consider Lipschitz functions wrt to d (the metric being lifted) and $d_{\mathbb{R}}$; second, we use $d_{\mathbb{R}}$ to compare the expected values of these functions.

For the purposes of bounding the refinement metric, it is more convenient to use a *quasimetric variant* of Kantorovich, in which $d_{\mathbb{R}}$ is replaced by $\mathbf{q}^{+}_{<}$.

Definition 4. *The Kantorovich quasimetric is the mapping* $\mathbb{K}^+ : \mathbb{M}\mathcal{A} \to \mathbb{M}\mathbb{D}\mathcal{A}$ *given by:*

$$\mathbb{K}^+(d)(\alpha, \alpha') \;:=\; \sup_{F:\,\mathbb{C}^{d,\mathbf{q}^+_<}\mathcal{A}} \mathbf{q}^+_<(\mathcal{E}_\alpha F, \mathcal{E}_{\alpha'} F) \;=\; \sup_{F:\,\mathbb{C}^{d,\mathbf{q}^+_<}\mathcal{A}} \mathcal{E}_{\alpha'} F - \mathcal{E}_\alpha F\ .$$

Note that the $\max$ in the definition of $\mathbf{q}^+_<$ is removed since the $\sup$ is anyway non-negative.

An important property of the Kantorovich metric is its *dual* formulation as the *Earth Mover's* distance $\mathbb{W}(d)$ (Def. 2, also known as Wasserstein metric). The well-known Kantorovich-Rubinstein theorem states that $\mathbb{K}(d) = \mathbb{W}(d)$ (when d is a proper metric and satisfies further assumptions, see below). In our case, however, we use $\mathbb{K}^+$ *instead of* $\mathbb{K}$, and moreover we apply it to distances d that are *not symmetric*. It turns out that the full Kantorovich-Rubinstein theorem still holds in our case (under the same conditions).

Theorem 5 (Kantorovich-Rubinstein, extended). *Let* $\mathcal{A}$ *be a Polish (i.e. complete and separable) metric space, let* d *be a lower semi-continuous distance function on* $\mathcal{A}$ *satisfying reflexivity and the triangle inequality. Then* $\mathbb{K}^+(d) = \mathbb{W}(d)$*.*

Proof. In [13, Thm. 1.3, page 19], a "Kantorovich duality" is proven, involving generic (lower semi-continuous) cost functions. Then, Kantorovich-Rubinstein is retrieved as a spacial case when the cost function is a metric [13, Thm. 1.14, page 34]. In the proof of the latter (which is quite short), only the reflexivity and the triangle inequality of d are used. Moreover, the change of $d_\mathbb{R}$ to $\mathbf{q}^+_<$ in the definition of $\mathbb{K}^+$ does not affect the proof. □

A surprising consequence of the above result is that, under the assumptions of Kantorovich-Rubinstein, we have:

$$\mathbb{K}^+(d) \;=\; \mathbb{K}(d) \;=\; \mathbb{W}(d)\ .$$

Note that the classes of Lipschitz functions in $\mathbb{K}^+$ an $\mathbb{K}$ are different, still maximizing over them turns out to give the same result! So, although the formulation of $\mathbb{K}^+$ is essential for computing capacity (see the next section), in fact it gives the standard Kantorovich construction.

A.2 Using Kantorovich/EMD to bound the refinement metric

The connection between Kantorovich and EMD allows us to obtain a bound on the refinement metric. To do so, we use the fact that posterior vulnerability, for any 1-bounded gain function, is non-expansive wrt convex separation quasimetric (Def. 3).

Lemma 1 ([5]). *For any* $g \in \mathbb{G}^{\updownarrow}\mathcal{X}$*,* V_g *is Lipschitz wrt* $\mathbf{q_{cs}}, d_\mathbb{R}$.

Since Kantorovich distance is given by maximizing over *any* Lipschitz function, it clearly gives an upper bound on the refinement metric. And the connection between Kantorovich and EMD allows us to efficiently compute this bound.

Theorem 2. *For any prior π and channels A, B it holds that*

$$\mathrm{ref}^{+}_{\pi,\mathbb{G}\ddagger}(\mathsf{A},\mathsf{B}) \quad \leq \quad \mathbb{W}(\mathbf{q_{cs}})([\pi \triangleright \mathsf{A}], [\pi \triangleright \mathsf{B}]) \ .$$

Proof. From Lemma 1 we get that $\mathrm{ref}^{+}_{\pi,\mathbb{G}\ddagger}(\mathsf{A},\mathsf{B}) \leq \mathbb{K}^{+}(\mathbf{q_{cs}})([\pi \triangleright \mathsf{A}], [\pi \triangleright \mathsf{B}])$, and the result follows from the (extended) Kantorovich-Rubinstein theorem.